

Partially-Distributed Resource Allocation in Small-Cell Networks

Sanam Sadr, *Student Member, IEEE*, and Raviraj S. Adve, *Senior Member, IEEE*

Abstract—We propose a four-stage hierarchical resource allocation scheme for the downlink of a large-scale small-cell network in the context of orthogonal frequency-division multiple access (OFDMA). Since interference limits the capabilities of such networks, resource allocation and interference management are crucial. However, obtaining the globally optimum resource allocation is exponentially complex and mathematically intractable. Here, we develop a *partially* decentralized algorithm to obtain an effective solution. The three major advantages of our work are: 1) as opposed to a fixed resource allocation, we consider load demand at each access point (AP) when allocating spectrum; 2) to prevent overloaded APs, our scheme is dynamic in the sense that as the users move from one AP to the other, so do the allocated resources, if necessary, and such considerations generally result in huge computational complexity, which brings us to the third advantage: 3) we tackle complexity by introducing a hierarchical scheme comprising four phases: user association, load estimation, interference management via graph coloring, and scheduling. We provide mathematical analysis for the first three steps modeling the user and AP locations as Poisson point processes. Finally, we provide results of numerical simulations to illustrate the efficacy of our scheme.

Index Terms—Small-cell networks, hierarchical resource allocation, Poisson point processes, graph coloring

I. INTRODUCTION

Each new generation of wireless communication systems promises to support a larger number of mobile users with higher data rates, new applications and ever-more stringent quality of service (QoS) requirements [1]. In a modern communication system where the mobile user is almost always connected to the network, supporting a large number of users with various applications results in a mix of traffic demands (bursty vs. continuous, high vs. low data rate) from the network point of view and high battery consumption from the user equipment (UE) point of view. Meeting these demands is getting harder largely due to the limited availability of transmission resources, most importantly wireless spectrum. This makes it impossible to completely separate concurrent transmissions in frequency, in turn, making interference the main factor that limits the capabilities of the wireless networks. Much of the relevant recent research focuses on the issue of interference e.g., [2], and network analysis in an interference-limited regime e.g., [3]–[6].

In the cellular context, an effective approach to increase capacity is to reduce the distance between the transmitter and

the receiver. This allows for a reduction in the transmit power and, in turn, improving the frequency reuse factor [7]. As an example of this approach, heterogeneous networks have been proposed to increase the area spectral efficiency¹ and the overall capacity through a revised network topology. The idea is to introduce a multi-tier network where each tier, also referred to as a layer, differs mainly in the density of its access points (APs) and the AP transmit power, hence coverage area. The traditional macro-cellular network would be one layer in the heterogeneous network with carefully planned deployment, the lowest density and the highest transmit power; the small-cell layer, on the other hand, is characterized by essentially random deployment, a much higher AP density and low AP transmit power. The randomness in the AP locations in small cells and their significantly greater number within a chosen geographical area precludes globally optimized resource planning. This necessitates new analysis techniques and algorithms beyond those for the centrally planned macro-cellular layer. In this paper, we focus on resource allocation techniques in small cells and propose a *partially*-distributed hierarchical scheme which can be applied to a *large-scale* network.

A. Related Work

The analysis of the signal-to-noise-plus-interference-ratio (SINR) in a heterogeneous network has been presented for a single-layer [3] and multi-layer [4]–[6] network using Poisson point processes (PPP). The presented analyses are based on the statistical distribution of the SINR in the network derived at a reference user randomly located in the cell. It is shown in [3] that in an interference-limited network, as is the case in small cells, the probability of coverage when the user is associated with a layer is independent of the AP density and the transmit power. The same result is shown in a multi-layer network with all the layers having the same layer association bias factor and path loss exponent [4]–[6]. While the coverage analysis here is a measure of the level of the received SINR, it is only synonymous with a chosen QoS if the reference user is guaranteed some resources. In other words, in a network with high density of users, while the users might be in coverage when considering the received SINR, the rate available to the user depends on the scheduling algorithms used at each AP.

Several resource allocation schemes have been proposed in the literature to mitigate RF interference in small cells in the context of femtocell networks. Femtocells are essentially user-deployed, indoor, small cells. Regardless of the details

This work was sponsored by TELUS and the National Science and Engineering Research Council (NSERC) Canada. The authors are with the Edward S. Rogers Sr. Department of Electrical and Computer Engineering, University of Toronto, 10 King's College Road, Toronto, Ontario, Canada M5S 3G4 (email: {ssadr, rsadve}@comm.utoronto.ca).

¹Area spectral efficiency is defined as the sum throughput per unit area that the system can provide per unit bandwidth.

of the system under consideration, they can be categorized into autonomous power control [8]–[12] and adaptive spectrum allocation [13]–[18]. The main feature of the first group is to adjust the AP coverage by setting the transmit power high enough to service its users but low enough so as not to interfere with the other APs on the same frequency of operation. The schemes in the second group manage interference by ensuring orthogonality between interfering APs.

A two-phase frequency assignment is proposed in [14], with a fixed, limited number of users per femtocell. Li et al. [15] viewed the user-deployed femtocells as the secondary system and the femtocell resource allocation as a cognitive spectrum reuse procedure. The idea is to adaptively adjust the channel reuse factor according to the location of the femtocell in the macrocell. Jointly optimizing power and spectrum, Kim et al. [17] proposed a scheme to maximize the total system capacity in dense networks. Treating the macro-users as primary users, the authors in [19] prioritize them by performing a hand-over to the nearby femtocell whenever the small-cell interference is high. The graph-based approach proposed in [18] maximizes the logarithmic average cell throughput to ensure proportional fairness among femtocells each serving a single user. A system level simulation of an open-access network was carried out by Claussen et al. [9], [10], and the obtained data rates at the reference users (one macro-user and one femto-user) were used to evaluate the system performance. Two main results are shown: 1) if autonomous power control is used by femtocells, adding APs has little impact on the macrocell throughput, and the impact is independent of the number of femtocells; 2) the total throughput significantly increases with the increase in the number of femtocell users, especially in the uplink. Similar results were reported in [4]–[6]. When analyzing the system, it is assumed that either the reference user is guaranteed some resources, e.g., in [2]–[6], or only voice is considered, e.g., in [16].

A more realistic simulation-based study of small-cell deployment in a heterogeneous network was reported by Coletti et al. [20]. The results suggest either coordination among layers or orthogonal spectrum allocation to improve outage rate. The authors of [21] propose a combination of fractional frequency reuse (FFR) and orthogonal spectrum allocation in a two-tier network differentiating between commercial and home-based femtocells.

An ambitious goal in dense networks is to achieve optimal but decentralized resource allocation. The problem of decentralized power allocation was first addressed by Foschini et al. [22]. They showed that there exists a fully distributed algorithm which requires only local information if there exists a common, known, SINR at which the system performance is globally optimum, and there exists a feasible but unknown power vector that achieves this SINR. Unfortunately, these assumptions are hard to satisfy in practice [23]. The proposed distributed algorithms in [24], [25] maximize the total system capacity ignoring user rate requirements and fairness among the users both within and among cells while [12] aims for proportional fairness ignoring individual user rate requirements. To obtain a distributed solution, the authors in [24], [25] simplify the network model to an “interference-ideal” network

where the total interference is constant and independent of user location in the cell.

B. Our Approach and Contributions

While there are several works on resource allocation in small-cell networks, as the literature survey above shows, it is hard to scale these algorithms to *large-scale* networks with multiple hundreds of nodes. Specifically, we consider the downlink of a large-scale network of small cells. Instead of focusing on SINR, we attempt to provide users with their desired data rate as they declare to their corresponding AP. Each AP transmits at its full transmit power and we focus on frequency allocation to avoid interference. To maintain fairness among users, we formulate the resource allocation in the form of a max-min normalized rate problem, maximizing the minimum ratio between the achieved and the desired rates. This problem is, in general, NP-hard.

The main contribution of this paper is an effective solution to this resource allocation problem with reasonable computational complexity. We propose a hierarchical scheme by decomposing the problem into four steps. This scheme has several advantages. As opposed to a fixed spectrum allocation as in [13] and [16], it considers the AP load, in terms of the number of users and their rate requirements, when allocating spectrum to the APs. This adaptivity in spectrum allocation allows for resources to follow user demands, i.e., high-rate users can be satisfied by a single AP. The proposed scheme is partially-distributed in the sense that three of the four steps are carried out locally and concurrently at each AP and, at worst, involve solving convex optimization problems using local information only. A single, graph coloring step must be executed at a central server. In contrast to the globally optimal solution requiring exponential complexity and global knowledge of channel state information, our hierarchical scheme imposes limited complexity and requires local knowledge only.

Given the difficulty with applying available algorithms to large-scale networks, we compare the results of our scheme with a network with fixed number of channels allocated to each AP, and show how load-awareness can effectively reduce the outage rate. As far as possible, we provide a mathematical analysis for the proposed scheme based on stochastic geometry [26]. In particular, we use independent homogeneous Poisson point processes for AP and user locations. PPPs have been shown to capture the inherent randomness in user and AP locations and yet provide tractable analyses compared to the grid-based models [3]. Necessarily requiring a number of simplifying assumptions, the analysis does provide results that track the related simulations and can, therefore, be used for system design.

The paper is structured as follows: in Section II, we describe the system model and formulate the resource allocation problem. The proposed hierarchical algorithm is presented in Section III with the outage analysis based on homogeneous PPP. This section also presents a complexity analysis. Section IV presents the simulation results in two parts: (a) comparing the performance of the proposed algorithm with its associated analysis; (b) comparing the performance with a fixed resource

allocation scheme. The parameters used in the simulations are based on the Long Term Evolution (LTE) standard. Section V concludes the paper with a discussion of the contributions.

II. DOWNLINK SYSTEM MODEL AND THE PROBLEM STATEMENT

Figure 1 illustrates the network under consideration. The model comprises K users and L APs distributed randomly in the network. The macro base station (BS) is provided an orthogonal frequency allocation, and our analysis considers only users connecting to the small cells. If the BS allocation is not orthogonal, it can be easily incorporated into the algorithm as another transmitting node in the network with its own power budget. Associated with each AP is its potential coverage area which depends on the environment and the transmit power. Due to the random geographical distribution of access points, the coverage area for some (if using the same time-frequency resource) may overlap. In other words, some users might be in a location covered by more than one AP and transmissions from these APs would interfere.

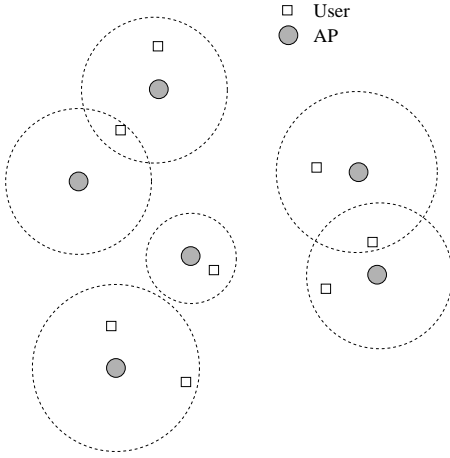


Fig. 1. Random distribution of APs and users in the network.

The optimization problem is formulated in the context of OFDMA as in the LTE standard. There are N effective frequency subchannels - physical resource blocks (PRBs) in LTE - available in the system each with a bandwidth of B . The channels between APs and users are modeled as frequency-selective Rayleigh fading with average power determined by distance attenuation and large scale fading statistics. The goal is to provide each user with its requested data rate. However, to achieve overall fairness in doing so, we formulate the problem to maximize the minimum normalized rate, i.e., max-min over all the users' achieved rates normalized by their requested data rates. The rate achieved on a specific channel is assumed to be given by its Shannon capacity; a gap function can be added to account for practical modulation and coding [27]. Each AP schedules its users in a manner to cancel the intra-cell interference. Therefore, the interference experienced by a user is due to the transmissions from all the access points, other than its own serving AP, that transmit on the same frequencies.

Under this setting, the general form of the resource allocation problem in the downlink is given by:

$$\begin{aligned} \max_{\{p_{k,n}^{(l)}\}, \{S_l\}} \min_k & \frac{1}{R_k} \left[\sum_{n=1}^N B \log_2 \left(1 + \frac{p_{k,n}^{(l)} h_{k,n}^{(l)}}{\sum_{i=1, i \neq l}^L p_{k,n}^{(i)} h_{k,n}^{(i)} + \sigma^2} \right) \right], \\ \text{subject to: } & \sum_{n=1}^N \sum_{k \in S_l} p_{k,n}^{(l)} \leq P_{tot}, \quad l = 1, 2, \dots, L, \\ & p_{k,n}^{(l)} \geq 0, \quad \forall k, n, l, \\ & S_i \cap S_j = \emptyset \quad i \neq j, \\ & \bigcup_{l=1}^L |S_l| = K, \end{aligned} \quad (1)$$

where $h_{k,n}^{(l)}$ and $p_{k,n}^{(l)}$ are, respectively, the channel power gain and the transmit power from AP l to user k on subchannel n . S_l is the set of users connected to and being serviced by AP l . R_k is the required rate by user k in bits per second (bps), σ^2 is the noise power, and B is the bandwidth of each subchannel. The sum, $\sum_{i=1, i \neq l}^L p_{k,n}^{(i)} h_{k,n}^{(i)}$, is the cumulative interference experienced by user k on subchannel n from all the access points except the serving AP indexed by l . The first constraint is on the total transmit power of each AP, while the second ensures non-negative transmit powers on each subchannel. The third constraint ensures that the sets, $\{S_l\}_{l=1}^L$, are disjoint, since each user is serviced by one and only one AP. The final constraint ensures that all the users in the system are scheduled by an access point. The sets, $\{S_l\}_{l=1}^L$, therefore, form a partition on the set of all users.

The objective of (1) is to find the optimal user associations, $\{S_l\}_{l=1}^L$, and power levels, $\{p_{k,n}^{(l)}\}_{l=1}^L$ determining which user should receive service from which AP on which subchannel, and how much power should be allocated to each subchannel. Being combinatorial, since it includes set selection, finding the optimal solution is exponentially complex. It seems infeasible from another point of view as well: it requires the knowledge of all the subchannels for all the users from all the APs at the central location. Getting this information to a central server would impose a huge overhead. Furthermore, this information needs to be updated every time the channel estimation is performed. Essentially, a resource allocation scheme based on global and perfect knowledge of SINR in a network of such scale is practically infeasible. This motivates developing partially-distributed, if suboptimal, solutions.

III. PROPOSED HIERARCHICAL ALGORITHM

In response to the infeasibility of obtaining the globally optimum solution, we propose a partially-distributed resource allocation scheme to decompose the problem into four steps:

- 1) Cell association: each user is associated with the AP that offers the highest long-term *average* received power (based, e.g., on a pilot and large-scale fading);
- 2) Load estimation: the load imposed by the users is *estimated* by each AP based on its users' data rate requirements and *average* channel gains;
- 3) Channel allocation: specific subchannels are allocated to APs based on coloring an interference graph;

- 4) Scheduling: each AP schedules its own users considering the users' required data rates and their *instantaneous* channel gains.

For each step, we derive the statistical distribution of the important quantities as accurately as possible, which are then used in analysis of the system performance, specifically the outage rate.

A. Partially-Distributed Resource Allocation

Step 1: Cell Association

The cell association is based on the large scale fading only. This implies that the cell association is the same whether the system considers the uplink or the downlink. At each user, the received power from all the APs are measured, and the user is associated with the AP that offers the highest long-term average received power. This ensures maximum data rate on average and makes the cell association independent of the instantaneous channel gains. This cell association is consistent with the downlink model for the system analysis based on PPP considered in [3] and [5].

Analysis: Considering only distance attenuation, each user's serving AP is the closest access point to the user. The statistical analysis of the connection distance is straightforward. Taking the user's location as the reference point, the connection distance is d if there is no other access point closer to the reference point. Modeling the AP locations by a homogeneous Poisson point process with density λ_f , the cumulative density function (CDF) of d can be written as:

$$\begin{aligned} F_D(d) &= \mathbb{P}(D < d) = 1 - \mathbb{P}(N_f(A_d) = 0) \\ &\stackrel{(a)}{=} 1 - \exp(-\lambda_f \pi d^2). \\ \Rightarrow f_D(d) &= 2\pi d \lambda_f \exp(-\lambda_f \pi d^2), \quad d \geq 0, \end{aligned} \quad (2)$$

where A_d is the area of a circle with radius d and the user at the center, $\mathbb{P}(B)$ denotes the probability of event B , and $N_f(A_d)$ is the number of APs in A_d . (a) results from the null probability of a 2-D Poisson process. Differentiating $F_D(d)$ with respect to d gives the probability density function (PDF) in the final equation.

Step 2: Load Estimation

With users having different rate demands, the objective at each AP is to estimate the minimum number of subchannels required to service its users. Each AP is aware of the requested rates and *instantaneous* channel gains for all the users that it serves. However, it does not know which subchannels it will be allocated, and fading is frequency selective. Therefore, it estimates its load using only the *average* channel gains. We emphasize that the load is defined here as the minimum frequency resources needed to meet the users' rate demands. We formulate this problem as a convex optimization problem

at each AP indexed by $l = 1, \dots, L$ given by:

$$\begin{aligned} \min_{n_k, P_k} \quad & \sum_{k \in S_l} n_k, \\ \text{subject to:} \quad & n_k B \log_2 \left(1 + \frac{P_k H_k}{n_k \sigma^2} \right) \geq R_k, \quad \forall k \in S_l, \\ & \sum_{k \in S_l} P_k \leq P_{tot}, \\ & P_k \geq 0, n_k \geq 0, \quad \forall k \in S_l. \end{aligned} \quad (3)$$

Here, n_k is the number (can be a fraction) of subchannels that AP l budgets for user $k \in S_l$. As before, S_l is the set of users supported by AP l . P_k and H_k are, respectively, the total power allocated to and the *average channel power* seen by user k . The first constraint ensures that the access point requests adequate resources to meet its users' demands. The objective is to minimize the total amount of spectrum needed by AP l . This is important since it affects the density of the interference graph in the next step.

Analysis: The load of the l -th AP in terms of the required amount of spectrum is a random variable given by:

$$N_l = \sum_{k=1}^m n_k, \quad (4)$$

where $m = |S_l|$ is itself a random variable representing the number of users connected to AP l . The CDF of N_l in the general form is given by:

$$F_{N_l}(n_l) = \mathbb{E}_{m, \{n_k\}} \left[\mathbb{P} \left(\sum_{k=1}^m n_k \leq n_l \right) \right], \quad (5)$$

where $\mathbb{E}_{m, \{n_k\}}$ denotes expectation with respect to m and $\{n_k\}_{k=1}^m$, and $n_k, k = 1, \dots, m$ are i.i.d random variables representing the required spectrum of each user. A thorough mathematical analysis of the load requires the knowledge of the number of users in each AP's coverage area, also referred to as *Voronoi* cells, and the users' distance to the serving AP. Due to the intractability of the problem, we derive the CDF for a special case where (i) all the APs have the same coverage area *on average*, (ii) equal power is allocated to each subchannel to estimate the load, and (iii) all the users connect at a distance equal to the most probable distance from the AP. We now state and prove the main result at this step.

Proposition: Under the assumptions stated above, the CDF of the AP load is given by:

$$F_{N_l}(n_l) = \mathcal{B}(n_l/n^*, K, p), \quad (6)$$

where $\mathcal{B}(n_l/n^*, K, p)$ is the CDF of a binomial distribution with K trials and probability of success $p = 1/L$ evaluated at n_l/n^* (n^* is defined later in (11) to be the number of subchannels required by each user).

Proof: In a network with L APs, the probability that a user is in the coverage area of a specific AP is $1/L$. Since user locations are random and independent, the probability mass function of the number of users m connecting to an access point is a binomial distribution given by:

$$P_u(m) = \binom{K}{m} p^m q^{K-m}, \quad (7)$$

where $p = 1/L$ and $q = 1 - p$.

From the first constraint in (3), the minimum required number of subchannels for user k with equal transmit power on each subchannel can be rewritten as:

$$n_k = \frac{R_k}{B \log_2 (1 + P_{\text{tot}} H_k / N \sigma^2)}, \quad (8)$$

where $H_k = L_0 d^{-\alpha}$, d is the distance between the AP and the user, L_0 is the path loss at the reference distance ($r_0 = 1\text{m}$), and α is the path loss exponent. Setting $\gamma_0 = P_{\text{tot}} L_0 / N \sigma^2$ and using $F_D(d)$ derived in (2), the CDF of the user load $F_{n_k}(n)$ can be written as:

$$\begin{aligned} F_{n_k}(n) &= \mathbb{P}(n_k < n) \\ &= \mathbb{P}\left(\frac{R_k}{B \log_2 (1 + \gamma_0 d^{-\alpha})} < n\right) \\ &= \mathbb{P}\left(\log_2 (1 + \gamma_0 d^{-\alpha}) > \frac{R_k}{nB}\right) \\ &= \mathbb{P}\left(d < \left(\frac{2^{R_k/nB} - 1}{\gamma_0}\right)^{-1/\alpha}\right) \\ &= F_D\left(\left(\frac{2^{R_k/nB} - 1}{\gamma_0}\right)^{-1/\alpha}\right). \end{aligned} \quad (9)$$

As shown in Fig. 5 in Section IV, this CDF approaches a step function as λ_f increases. In other words, in a dense network of small cells, the variance of the user load decreases. This motivates one to represent the user load in high-density AP networks with a single number rather than a random variable. Amongst various possibilities, the *most probable link distance* best predicts this number.

Differentiating $f_D(d)$ with respect to d and setting it to zero, the most probable link distance is given by:

$$d^* = \sqrt{\frac{1}{2\pi\lambda_f}}. \quad (10)$$

Inserting d^* in (8), the number of subchannels required by each user is given by:

$$n^* = \frac{R_k}{B \log_2 (1 + \gamma_0 (d^*)^{-\alpha})}. \quad (11)$$

Using (4), the CDF of the AP load is then derived as:

$$\begin{aligned} F_{N_l}(n_l) &= \mathbb{P}(N_l \leq n_l) = \mathbb{P}(n^* m \leq n_l) \\ &= \mathbb{P}(m \leq n_l / n^*) \\ &= \mathcal{B}(n_l / n^*, K, p), \end{aligned} \quad (12)$$

where the last equation results from the distribution of m derived in (7) and the proof is complete. ■

For a large K , the binomial distribution is very well approximated by a Poisson distribution with parameter $\eta = K/L$ when η is small, and by a normal distribution when η is large [28]. In this paper, we will make use of the Poisson approximation.

Step 3: Channel Allocation Among APs Using Graph Coloring

After steps 1 and 2, users have been assigned to APs and the APs have estimated their loads. We now come to the crucial step of allocating subchannels to APs. Specifically,

the objective at this step is to allocate the spectrum to the APs considering their load and the interference they can potentially cause to other small cells. In this paper, we consider a resource allocation scheme which avoids interference. To do so, we ensure that two neighboring small cells do not use the same frequencies. Small cells are considered neighbors if they potentially interfere with each other, i.e., their potential coverage areas overlap. Unlike the previous two steps (and the next step), this allocation is *centralized*.

All the access points report their load $\{N_l\}_{l=1}^L$ to the central server. Ideally, the central server should allocate an orthogonal set of subchannels to every AP that also meets its users' requirements. Given N total available subchannels, if $N \geq \sum_{l=1}^L N_l$, each AP is easily satisfied. Realistically, however, this is highly unlikely; hence, the server must reuse channels across multiple APs. This can cause interference, and so the allocation must ensure that the interfering APs (APs with overlapping coverage areas) are assigned orthogonal frequency resources. As a consequence, it may be that all APs' load demands cannot be satisfied. Alternatively, the goal is to assign subchannels to APs *proportional* to their estimated load while eliminating the interference among them. To do so, we use graph coloring by the central server.

Graph algorithms have been used as a tool for channel assignment in multi-cellular networks, e.g., in [29]–[34], with the nodes representing either access points or users. Chang et al. [30] formulated the spectrum allocation in a macrocellular network in the form of max \mathcal{K} -Cut with a fixed number of channels (or colors). Each node in the graph corresponds to a mobile device or user. The interference among users is denoted by weighted edges taking into account not only the distance between the users but also the anchor (serving) and the neighboring base stations. The objective is to partition the users into \mathcal{K} clusters with maximum inter-cluster weight. This technique allows for asymmetrical channel allocations among the base stations. Authors in [32] proposed a two-step graph coloring approach for multicell OFDMA networks in which the users are clustered in a manner to minimize the total number of colors based on geographic user locations. In the second step, the subchannels are allocated based on instantaneous channel conditions. In graph-based schemes, wherever users correspond to graph nodes as in [32] and [34], user mobility results in rapid changes of the interference graph. Since we deal with small cells in a dense network, this computation is added to the signalling overhead due to hand-off and synchronization among APs making it impractical. The authors in [31] differentiate between the cell centre and cell edge hence allowing for FFR. This approach assumes large cells, an assumption that is not valid here. Finally, the two-step spectrum allocation algorithm proposed in [33] uses the instantaneous channel information in deriving femtocells' utilities while coloring the graph resulting in increased complexity and signalling overhead.

In our approach, the nodes of the graph represent access points. An edge connects two nodes if they potentially interfere based on *large-scale* statistics. We make this choice to ensure that the graph does not change rapidly with each channel realization. While here we use an unweighted graph, this is not

fundamental to the proposed scheme. A weighted graph can very well be used instead, at the cost of increased complexity as long as the edge weights correctly reflect the intensity of the interference between any two nodes. Since each color corresponds to a single subchannel, to account for the AP loads, we modify the interference graph as follows: as opposed to the conventional approach, AP l is represented by not one but $\lceil N_l \rceil$ nodes forming a complete subgraph ($\lceil \cdot \rceil$ denotes the “ceiling” function). The problem of channel assignment among APs becomes a graph coloring problem where two interfering nodes (nodes connected with an edge) should not be assigned the same color. An example of a three AP network with (estimated) $N_1 = 1, N_2 = N_3 = 3$ is illustrated in Fig. 2. The corresponding interference graph is shown in Fig. 3. AP #1 potentially interferes with AP #2 and AP #3. So, allocations to AP #1 cannot be reused for AP #2 or 3. However, since AP #2 does not interfere with AP #3, frequencies can be reused across these two APs.

One solution to the coloring problem is illustrated in Fig. 4. It is worth noting that the coloring is not unique. For example, a simple index shift (a re-ordering of the association between the graph and the frequency slots) is an equally valid solution to the graph coloring problem. Amongst these many solutions, there is one optimal solution that best meets the demands of the individual APs based on the specific instantaneous realizations of user-AP channels. However, none of this information is available at the central sever; this lack of optimality is the penalty for using a distributed algorithm with limited knowledge of CSI.

For arbitrary graphs, graph coloring is an NP-hard problem. The optimal coloring is possible with low complexity algorithms if the interference graph is sparse such that each node is connected to at most N nodes where N is the total number of available colors. Such graphs can be colored with a modified Breadth First Search (BFS) algorithm with complexity of $\mathcal{O}(|V| + |E|)$ with $|E| = \mathcal{O}(|V|)$ where $|E|$ and $|V|$ are the cardinality of edges and vertices respectively. We adopt the heuristic (greedy) algorithm proposed by Br elaz [35]: at every iteration, the vertex which is adjacent to the greatest number of differently-colored neighbours is colored, with a new color if necessary (until colors are exhausted). A major advantage of our proposed hierarchical scheme is that by carrying out the graph coloring step in a distributed manner as proposed in [36], we achieve a fully distributed scheme.

Outage Analysis: While coloring an interference graph results in a higher spatial reuse factor for a channel, outage is inevitable if an access point and its interfering neighbours require more than the available number of subchannels. Let \tilde{p} be the AP pilot power and τ be its receive threshold determining the coverage area. Two APs separated by $\tilde{r} = 2\tilde{d}$ interfere if the received power at the midpoint of the distance between the two satisfies $\tilde{p}\tilde{d}^{-\alpha} > \tau$. Now, taking a randomly chosen AP as the reference, we prove the following statement.

Proposition: The probability that the reference AP and its neighbours will not have enough spectrum, hence leading to

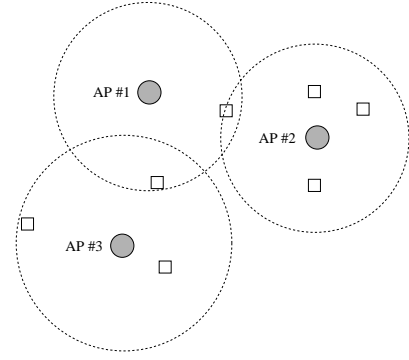


Fig. 2. A network of 3 APs. $N_1 = 1, N_2 = N_3 = 3$.

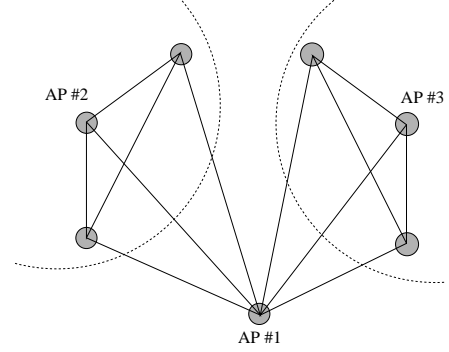


Fig. 3. Interference graph corresponding to Fig. 2.

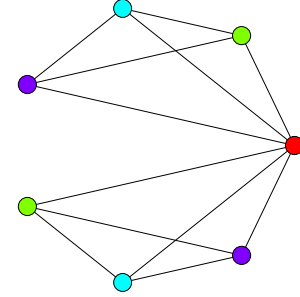


Fig. 4. Graph coloring corresponding to Fig. 3. Minimum number of colors is 4, with both optimal and suboptimal coloring algorithms.

an outage, is given by:

$$P_o = 1 - \sum_{\tilde{L}=1}^L \left[\text{Pois}(N/n^*, \tilde{L}\eta) \mathbb{P} \left(N_f(A_{\tilde{r}}) = \tilde{L} \right) \right], \quad (13)$$

where \tilde{L} is the number of APs in the area $A_{\tilde{r}} = \pi\tilde{r}^2$, $\mathbb{P} \left(N_f(A_{\tilde{r}}) = \tilde{L} \right) = \frac{e^{-\lambda_f A_{\tilde{r}}}}{\tilde{L}!} (\lambda_f A_{\tilde{r}})^{\tilde{L}}$, and $\text{Pois}(N/n^*, \tilde{L}\eta)$ is the CDF of the Poisson random variable with mean $\tilde{L}\eta$ evaluated at N/n^* .

Proof: Let $\tilde{L} - 1$ be the total number of APs interfering with the reference AP, i.e., \tilde{L} APs cannot share the same subchannel. The CDF of the minimum required number of subchannels \tilde{N} such that our reference AP can support its

users (without interfering with its neighbours) is given by:

$$\begin{aligned} F_{\tilde{N}}(\tilde{n}|\tilde{L}) &= \mathbb{P}(\tilde{N} \leq \tilde{n}|\tilde{L}) \\ &= \mathbb{P}\left(\sum_{l=1}^{\tilde{L}} N_l \leq \tilde{n}|\tilde{L}\right) \\ &= \text{Pois}(\tilde{n}/n^*, \tilde{L}\eta), \end{aligned} \quad (14)$$

where we use the fact that the sum of \tilde{L} independent Poisson random variables with mean η is a Poisson random variable with mean $\tilde{L}\eta$. The probability of outage is then given by:

$$\begin{aligned} P_o &= \mathbb{E}_{\tilde{L}} \left[\mathbb{P}(\tilde{N} > N) \right] = \mathbb{E}_{\tilde{L}} [1 - F_{\tilde{N}}(N)] \\ &= \sum_{\tilde{L}=1}^L \left[1 - F_{\tilde{N}}(N|\tilde{L}) \right] \mathbb{P}(N_f(A_{\tilde{r}}) = \tilde{L}) \\ &= 1 - \sum_{\tilde{L}=1}^L \left[\text{Pois}(N/n^*, \tilde{L}\eta) \mathbb{P}(N_f(A_{\tilde{r}}) = \tilde{L}) \right], \end{aligned} \quad (15)$$

and the proof is complete. ■

Unfortunately, it does not appear that this final expression can be further simplified. However, the summation is easily evaluated numerically. In a system with L points of a Poisson process (representing APs) uniformly distributed, \tilde{L} is binomial with L trials and probability of success $(\tilde{r}/R_c)^2$ where R_c is the radius of the macrocell under consideration. $F_{\tilde{N}}(\tilde{n})$ is the CDF of a Poisson random variable and can be calculated using the *incomplete gamma function*.

Step 4: Resource Allocation Among Users

At the end of step 3, each AP is assigned an integer number of subchannels without interfering with its neighbours. The problem at each AP is now reduced to maximizing the minimum rate of the users it services relative to their requested data rate. In doing so, the AP considers the *instantaneous* CSI of the subchannels it has been assigned. In this regard, it is worth restating that the previous three steps were based on *average* channel powers.

Let \tilde{N}_l be the number of subchannels assigned to AP l in Step 3; this is not necessarily equal to its estimated requirement N_l . The scheduling problem at each AP is formulated as:

$$\begin{aligned} \max_{p_{k,n}, c_{k,n}} \min_k & \frac{1}{R_k} \left[B \sum_{n=1}^{\tilde{N}_l} c_{k,n} \log_2 \left(1 + \frac{p_{k,n} h_{k,n}}{c_{k,n} \sigma^2} \right) \right] \\ \text{subject to} & \sum_{n=1}^{\tilde{N}_l} \sum_{k \in S_l} p_{k,n} \leq P_{tot}, \quad p_{k,n} \geq 0 \\ & \sum_{k \in S_l} c_{k,n} = 1, \\ & c_{k,n} \geq 0 \quad \forall n, \quad k \in S_l \end{aligned} \quad (16)$$

where $c_{k,n}$ is the fraction of subchannel n allocated to user k . $h_{k,n}$ and $p_{k,n}$ are the channel power gain and the transmit power to user k on subchannel n . This is a standard convex optimization problem. An even simpler alternative is to divide power equally amongst the \tilde{N}_l subchannels, $p_{k,n} = P_{tot}/\tilde{N}_l$, leading to a linear program in which users share the resources

using time-division².

B. Complexity Analysis

Step 1 - cell association: Each user connects to the AP with the highest average received power. Finding the AP with the maximum received power requires L comparisons at each user. Hence, the complexity of this step is of the order $\mathcal{O}(L)$ for each of K users.

Step 2 - load estimation: This is a convex optimization problem with the complexity depending on the solution method, e.g., the interior-point method or Newton-Raphson. Furthermore, the number of iterations in each depends on the stopping criterion. In Newton-Raphson method, the computational complexity mainly results from finding the update direction. It is shown that the computational complexity of each iteration is $\mathcal{O}(M^3)$, where M is the number of users connected to one AP. The details are provided in the Appendix.

Step 3 - spectrum allocation among APs: This step consists of two smaller steps. 1) Forming the interference graph: any two APs closer than a threshold distance are connected with an edge. Hence, the complexity of this step is of the order $\mathcal{O}(L^2)$. 2) Graph coloring: the complexity depends on the density of the graph algorithm as provided in Step 3 of Subsection III-A. Since this step is carried out at the central unit, with slower changes compared to locally solved problems, more sophisticated algorithms can be used.

Step 4 - scheduling: The normalized rate scheduling at this step is a modified version of the problem formulated by Rhee et al. [37]. A special case is when equal transmit power is used on all the subchannels leading to close to optimum performance when the system benefits from user-channel diversity. The proposed suboptimal subchannel allocation with equal transmit power has complexity of $\mathcal{O}(M * \tilde{N}_l)$, where \tilde{N}_l is the number of PRBs allocated to the AP.

IV. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed scheme and the mathematical analysis presented in the previous section. The simulations are based on the LTE standard closely following [38]. The downlink transmission scheme for an LTE system is based on OFDMA where the available spectrum is divided into multiple subcarriers each with a bandwidth of 15kHz. Resources are allocated to users in blocks of 12 subcarriers referred to as physical resource blocks; hence, the bandwidth of each PRB is 180kHz and is used as the signal bandwidth in calculating the noise power. The receiver noise power spectral density is set to -174dBm/Hz with an additional noise figure of 9dB at the receiver. Here, we consider the maximum LTE bandwidth (20MHz); $N = 50$ PRBs are allocated to the small-cell network. The APs are distributed within a circle of radius 100m, i.e., covering $3.14 \times 10^4 \text{ m}^2$. Table I lists the parameters used in all the simulations, unless otherwise specified.

²Standards such as LTE provide the ability to reassign physical resource blocks every millisecond. Such flexibility is reflected here as time-sharing of the subchannels.

TABLE I
SIMULATION PARAMETERS

Parameters	Value
Carrier frequency	2 GHz
Channel bandwidth	20 MHz
Carrier spacing	15 kHz
Resource block (B)	180 kHz
Number of PRBs available (N)	50
Transmit power	20dBm
Antenna gain	0dB
Antenna configuration	1×1
Noise Figure in UE	9dB
Minimum distance of user Penetration loss (wall/window)	1m from AP 10dB/3dB
\bar{d}	20m
Region covered (R_c)	Circle of radius 100m

The path loss between the access point and the user accounts for indoor and outdoor propagation:

$$PL = 38.46 + 20 \log_{10}(d_{in}) + 37.6 \log_{10}(d) + L_p + L_s \quad (17)$$

where d_{in} is the distance between the access point and the external wall or window and has a uniform distribution between 1m and 5m; L_p is the penetration loss and is set to 10dB or 3dB (with equal probability) representing an external wall and window respectively; L_s accounts for shadowing and is modeled by a log-normal random variable with standard deviation of 10dB. Finally, assuming Rayleigh fading, the instantaneous power of the received signal is modeled as an exponential random variable with the mean equal to the average received power [39]. In the mathematical analysis, the path loss exponent, α , is set to 3. The multipath environment is such that the fading is effectively flat for the 12 subcarriers in one PRB but rich enough to yield an independent fade on each PRB. Each PRB is then allocated to a user for a *subframe* duration of 1ms.

A. Validating the Mathematical Analysis

First, we validate the mathematical analysis of the proposed hierarchical algorithm. Figure 5 plots the CDF of the user load $F_{n_k}(\cdot)$ for two cases: high ($\lambda_f = 1/100$ or 1 AP per $100m^2$ corresponding to 314 APs) and low ($\lambda_f = 1/1000$) AP density. At each run of the simulation, the APs and the users are randomly located in the cell according to the given densities with user density denoted by λ_u . Each user connects to the closest access point. The user load is then calculated for a randomly chosen user (as the reference) with equal transmit power on all the PRBs considering only the distance attenuation. The result is compared to the CDF derived in (9). As is clear, the analysis matches the simulation results exactly. Also, as expected, the CDF is shifted to the right in a network with lower AP density; the reason is a larger connection distance and a higher load (measured in terms of required subchannels) imposed by the user for the same data rate. Crucially, the figure shows that the CDF of the individual

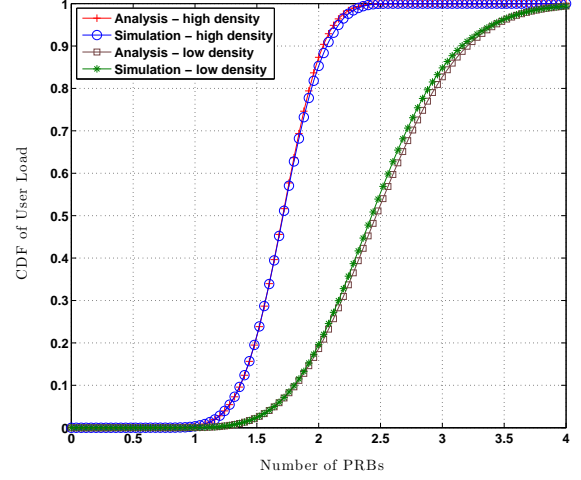


Fig. 5. CDF of the user load $F_{n_k}(\cdot)$ for high ($\lambda_f = 1/100m^2$) and low ($\lambda_f = 1/1000m^2$) AP density. $R_k = 5$ Mbps.

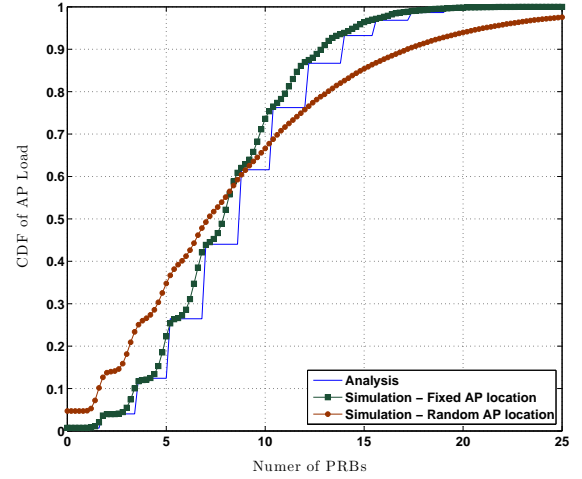


Fig. 6. CDF of the AP load $F_{N_l}(\cdot)$. $\lambda_f = 1/(100m^2)$, $\lambda_u = 5\lambda_f$ and $R_k = 5$ Mbps.

user load approaches a step function at a higher AP density. This justifies the approximation in (11).

Figure 6 plots the CDF of the load of a randomly chosen access point in two scenarios: (i) random AP location as in PPP; (ii) fixed AP location with equal coverage area for all the APs. In both cases, the load of the reference AP is the sum of its users' load. The CDF derived in (12) slightly deviates from the simulation results due to two main simplifying assumptions: 1) all APs have the same coverage area on average leading to a binomial distribution for the number of users connecting to an AP; 2) the most probable connection distance has been assumed for all the users connecting to an AP based on the AP density in the system. In a system where APs are randomly located, although uniformly distributed, they might have different coverage areas. For the purpose of comparison, a network with fixed AP locations is considered with equal coverage area for all APs. In such a system, $P_u(m)$

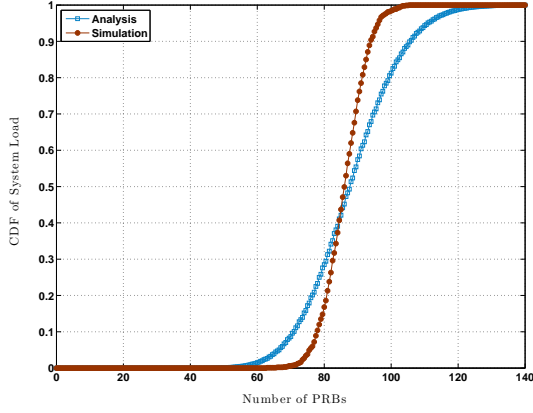


Fig. 7. CDF of the system load $F_{\tilde{N}}(\cdot)$. $\lambda_f = 1/(100m^2)$, $\lambda_u = 5\lambda_f$ and $R_k = 1\text{Mbps}$.

follows the binomial distribution. A small deviation still exists due to the second simplifying assumption.

The CDF of the system load, $F_{\tilde{N}}(\cdot)$ is shown in Fig. 7. The system load is the sum of the total number of PRBs required by a randomly chosen access point (as the reference) and all its interfering APs derived in (14). If the distance between two APs is no more than $2\tilde{d}$, where \tilde{d} is the radius of the coverage circle of one AP, the pair are assumed to interfere with each other. A lower SNR threshold results in a larger coverage area for each AP and a denser interference graph. Note that, for any SNR threshold, this is the worst-case scenario assuming the user is in the midpoint of the distance between the two APs. In practice, whether two APs interfere can be estimated more accurately by each AP based on a pre-defined coalition threshold [40] and reported to the central server (most protocols allow an access point to keep a “neighbour” list). Figure 8 plots the corresponding outage probability as a function of the common user demand $R_k = R$. Again, the analysis captures the essential behaviour of the outage probability with a slight mismatch due to the simplifying assumptions mentioned before for tractability.

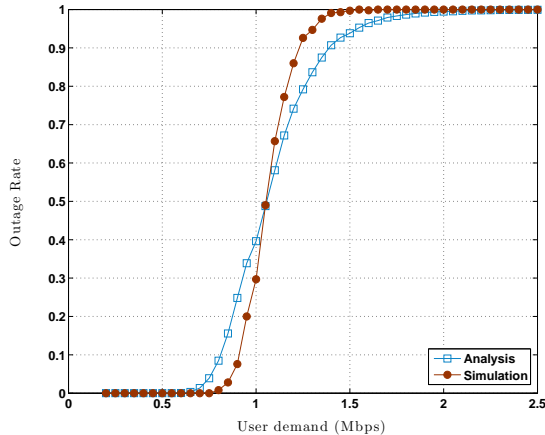


Fig. 8. Outage rate versus user demand. $\lambda_f = 1/(200m^2)$ and $\lambda_u = 5\lambda_f$.

B. Performance Comparison

In this section, we illustrate the performance of the proposed hierarchical scheme and compare its performance with a fixed-allocation scheme. The globally optimal solution through exhaustive search is impossible to obtain in a reasonable time due to its exponential complexity and so is not compared to.

The fixed-allocation scheme is as follows: each AP is assigned N_{AP} PRBs randomly chosen out of the N PRBs available to the small-cell network. The cell association and user level scheduling is the same for both algorithms. Hence, the main differences between the fixed-allocation and the proposed hierarchical scheme are the element of interference management and the effect of load estimation in dynamic distribution of PRBs among APs. The purpose of such dynamic distribution is to improve the user’s achieved rate in the whole system proportional to its demand. A user is considered to be in outage when it receives less than its required data rate.

In a network with fixed spectrum allocation, N_{AP} affects the density of the interfering APs [3]. We first consider the performance of the fixed-allocation scheme as a function of N_{AP} , the number of PRBs assigned to each AP. Figure 9 plots the number of the users in outage normalized by the total number of the users for two user densities. All the users request the same data rate of 1.5Mbps. As shown in the figure, the outage decreases with N_{AP} to a point where it is saturated such that further increase in N_{AP} results in higher interference and hence, outage. In this example, $N_{AP} = 18$ gives the best performance for the given AP and user densities. In subsequent testing, we use a fixed value of $N_{AP} = 18$. This allows for a comparison of our results to the *best-case scenario* for the fixed-allocation scheme.

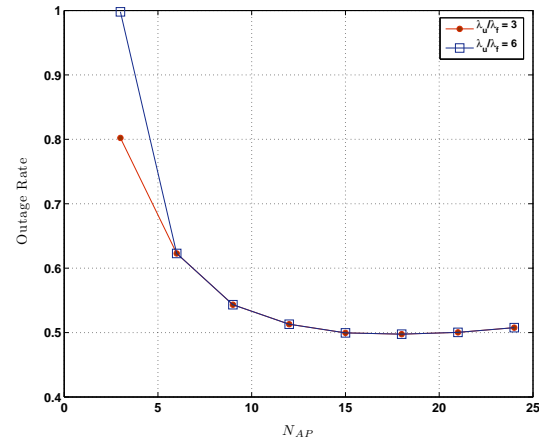


Fig. 9. Outage rate as a function of the fixed number of PRBs assigned to each AP. $\lambda_f = 1/(200m^2)$, $\lambda_u = 3\lambda_f$ and $\lambda_u = 6\lambda_f$. $R_k = 1.5\text{Mbps}$.

The outage (in a log-scale) for both schemes versus the user demand is shown in Fig. 10. As expected for both algorithms, the number of users in outage increases with the increase in the user demand. In both user-to-AP densities (λ_u/λ_f), there is an obvious gain with using the hierarchical scheme - the outage rate improves by up to an order of magnitude at the lower user demands. It is worth noting that at high user

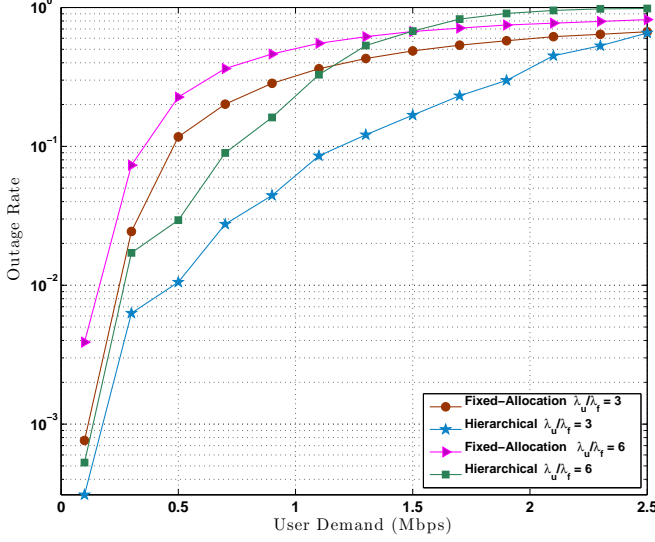


Fig. 10. Users at outage in both schemes versus the user demand. $\lambda_f = 1/(200m^2)$, $\lambda_u = 3\lambda_f$ and $\lambda_u = 6\lambda_f$. $N_{AP} = 18$ for the fixed-allocation.

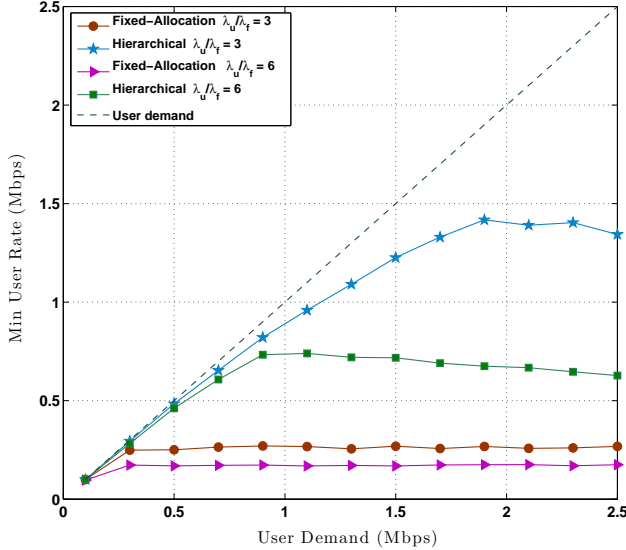


Fig. 11. Average minimum user achieved rate for both schemes versus the user demand. $\lambda_f = 1/(200m^2)$, $\lambda_u = 3\lambda_f$ and $\lambda_u = 6\lambda_f$. $N_{AP} = 18$ for the fixed-allocation.

demands (above $R_k = 1.5\text{Mbps}$ for $\lambda_u/\lambda_f = 6$ and above $R_k = 2.5\text{Mbps}$ for $\lambda_u/\lambda_f = 3$) the fixed-allocation scheme actually has a lower outage. This is to be expected since in the hierarchical scheme, any two APs that are less than \tilde{r} meters apart are connected in the interference graph regardless of the degree of interference. This results in higher system load estimation and smaller number of PRBs allocated to *each* AP in the system. In the fixed-allocation scheme on the other hand, due to the lack of any interference management, the effect of concurrent transmissions are added exactly according to

the path loss model. Hence, the comparison here is between the worst-case scenario of the hierarchical scheme and the best-case scenario of the fixed-allocation. Using weighted interference graphs and more sophisticated graph algorithms in Step 3 should improve the performance at the cost of increased computational complexity.

A significant advantage of the proposed scheme is to shift the available spectrum from the underloaded APs to the overloaded APs to achieve higher level of fairness over all the users in the network. Examining the minimum achieved user rate in the system in Fig. 11 shows that our proposed scheme achieves this goal. While both schemes converge to a constant value with the increase in the user demand, the hierarchical scheme reaches a higher value (more than twice the minimum rate in fixed-allocation) for both user densities. Any user achieved rate would fall between the minimum achieved rate and the user demand (the maximum rate assigned to the user represented by the dotted line). The closer the two are, the higher level of fairness is achieved. In other words, the hierarchical scheme achieves a higher degree of fairness and is more efficient in terms of allocating resources in comparison to the fixed resource allocation.

It is worth noting that this gain is higher in a system with a lower user-to-AP density. The reason is that in a network with independent user locations, the probability density function of the user distribution in a unit area (and hence the AP load) approaches a dampened normal distribution as opposed to a Poisson distribution with the increase in the user density. This effect corresponds to a smaller variance in the AP load in the system. The proposed scheme is most effective in systems with higher possibility of underloaded and overloaded APs existing at the same time which explains the higher gain in $\lambda_u = 3\lambda_f$ compared to $\lambda_u = 6\lambda_f$.

As a final comparison, Fig. 12 plots the total throughput of the system. The higher throughput in the hierarchical scheme is the result of higher user achieved rate as discussed above.

V. CONCLUSION

In this paper, we have proposed a hierarchical 4-stage resource allocation scheme for large-scale small-cell networks. The main advantage of the proposed scheme is decomposing a complex non-convex optimization problem into several smaller convex problems with smaller sets of optimization variables. The result is a low complexity scheme effective with a large problem size; in our simulations, the resource allocation can be achieved across as many as 314 APs.

The rationale behind the introduced hierarchy is as follows: user locations combined with various user demands result in a non-uniform distribution of the load in the system. Access points will experience very different load demands as shown in Fig. 6. Hence, in an efficient allocation, resources should be dynamically allocated to meet this load. Here, this load is approximated at each AP by solving the related optimization problem based on the local information, i.e., users' demand and the average channel power. To do so, the APs do not require any global information. Load estimation and the last step of resource allocation at the APs would, in practice, be

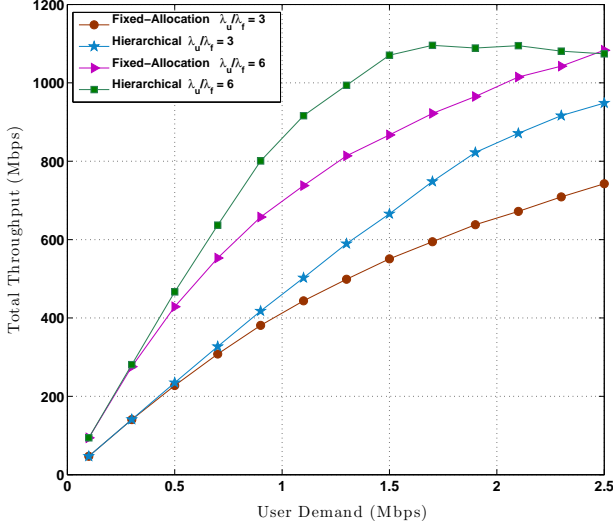


Fig. 12. Total throughput of the system for both schemes versus the user demand. $\lambda_f = 1/(200m^2)$, $\lambda_u = 3\lambda_f$ and $\lambda_u = 6\lambda_f$. $N_{AP} = 18$ for the fixed-allocation.

solved in parallel at each AP. Only a single step of graph coloring is executed at a central server - this centralized step ensures orthogonal allocations to APs that interfere with each other.

The central server only requires knowledge of the demands made by each AP which significantly reduces the signaling overhead. The results confirmed an increase in the user's achieved data rate with the proposed hierarchical scheme as apposed to the fixed resource allocation. Across a wide range of user rate demands, the scheme results in a significantly lower outage. Finally, the simulations confirm that the proposed scheme is most effective in systems with a higher possibility of underloaded and overloaded APs existing simultaneously.

APPENDIX

COMPLEXITY ANALYSIS FOR STEP 2: LOAD ESTIMATION

The load estimation problem in (3) is equivalent to finding the minimum of the following cost function

$$\mathcal{L} = \sum_{k \in S_l} n_k + \sum_{k \in S_l} \mu_k \left[R_k - n_k B \log_2 \left(1 + \frac{P_k H_k}{n_k \sigma^2} \right) \right] + \mu_0 (\sum_{k \in S_l} P_k - P_{tot}), \quad (18)$$

where $\{\mu_i\}_{i=0}^{|S_l|}$ are the Lagrangian multipliers. Differentiating (18) with respect to P_k and n_k , and setting each derivative to 0, we obtain:

$$\frac{\partial \mathcal{L}}{\partial n_k} = 1 - \mu_k C \left[\ln \left(1 + \frac{P_k H_k}{n_k \sigma^2} \right) - \frac{P_k H_k}{\sigma^2 n_k \left(1 + \frac{P_k H_k}{n_k \sigma^2} \right)} \right] = 0, \quad (19)$$

$$\frac{\partial \mathcal{L}}{\partial P_k} = -\frac{\mu_k C \frac{H_k}{\sigma^2}}{1 + \frac{P_k H_k}{\sigma^2 n_k}} + \mu_0 = 0, \quad (20)$$

for $k = 1, 2, \dots, M$, where $C = B \log_2 e$ and $M = |S_l|$ are for simplicity of presentation. From (20), we obtain:

$$\frac{\mu_k H_k}{1 + \frac{P_k H_k}{n_k \sigma^2}} = \frac{\mu_1 H_1}{1 + \frac{P_1 H_1}{n_1 \sigma^2}}. \quad (21)$$

Combined with the power constraint $\sum_{k \in S_l} P_k = P_{tot}$, and the rate constraints $n_k B \log_2 \left(1 + \frac{P_k H_k}{n_k \sigma^2} \right) = R_k$, $k = 1, 2, \dots, M$, there are $3M$ variables $\{P_k, n_k, \mu_k\}_{k=1}^M$ in the set of $3M$ non-linear equations in (19)-(21). Iterative methods such as Newton-Raphson can be used to obtain the solution, with the complexity mainly due to finding the update direction.

Denote $\mathbf{X} = [P_1, \dots, P_M, n_1, \dots, n_M, \mu_1, \dots, \mu_M]^T$ as the variables and $\mathbf{G} = 0$ as the square system of non-linear equations. The update direction $\Delta \mathbf{X}$ is found solving the following equation:

$$\mathcal{J}(\mathbf{X}) \Delta \mathbf{X} = -\mathbf{G}(\mathbf{X}), \quad (22)$$

where $\mathcal{J}(\mathbf{X})$ is the Jacobian matrix of $\mathbf{G}(\mathbf{X})$ evaluated at \mathbf{X} . Using Gauss-Jordan method, the complexity of the algorithm to solve for $\Delta \mathbf{X}$ in each iteration is of the order $\mathcal{O}(M^3)$. A special case is when equal transmit power is used on the subchannels; in this case, estimating the load at each AP has the complexity of the order $\mathcal{O}(M)$ (due to M divisions at each AP).

REFERENCES

- [1] D. C. Kilper, G. Atkinson, S. K. Korotky, S. Goyal, P. Vetter, D. Suvakovic, and O. Blume, "Power trends in communication networks," *IEEE J. Sel. Topics Quantum Electron.*, vol. 17, no. 2, pp. 275–284, Mar-Apr. 2011.
- [2] M. Haenggi and R. K. Ganti, "Interference in large wireless networks," *Foundations and Trends in Networking*, vol. 3, no. 2, pp. 127–248, 2008.
- [3] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, Nov. 2011.
- [4] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "A tractable framework for coverage and outage in heterogeneous cellular networks," *Inf. Theory and Appl. Workshop (ITA)*, Feb. 2011.
- [5] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Trans. Wireless Commun.*, vol. 11, no. 10, pp. 3484–3495, Oct. 2012.
- [6] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.
- [7] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [8] A. Barbieri, A. Damnjanovic, T. Ji, J. Montojo, Y. Wei, D. Malladi, O. Song, and G. Horn, "LTE femtocells: System design and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 586–594, Apr. 2012.
- [9] H. Claussen, "Performance of macro- and co-channel femtocells in a hierarchical cell structure," in *Proc. PIMRC*, Sept. 2007.
- [10] H. Claussen, L. T. W. Ho, and L. G. Samuel, "An overview of the femtocell concept," *Bell Labs Technical Journal - Wiley*, vol. 13, no. 1, pp. 221–245, 2008.
- [11] V. Chandrasekhar, J. G. Andrews, T. Muharemovict, Z. Shen, and A. Gatherer, "Power control in two-tier femtocell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4316–4328, Aug. 2009.
- [12] F. Ahmed, A. A. Dowhuszko, and O. Tirkkonen, "Distributed algorithm for downlink resource allocation in multicarrier small cell networks," in *Proc. ICC*, pp. 6802–6808, Jun. 2012.
- [13] V. Chandrasekhar and J. G. Andrews, "Spectrum allocation in two-tier networks," in *Proc. IEEE Asilomar Conf. on Signals, Systems and Computers*, pp. 1583–1587, Oct. 2008.

- [14] D. Lopez-Perez, A. Ladanyi, A. Juttner, and J. Zhang, "OFDMA femtocells: A self-organizing approach for frequency assignment," in *Proc. PIMRC*, pp. 2202–2207, Sept. 2009.
- [15] Y.-Y. Li, M. Macuha, E. S. Sousa, T. Sato, and M. Nanri, "Cognitive interference management in 3G femtocells," in *Proc. PIMRC*, pp. 1118–1122, Sept. 2009.
- [16] Y. Zhang, "Resource sharing of completely closed access in femtocell networks," in *Proc. WCNC*, Apr. 2010.
- [17] J. Kim and D.-H. Cho, "A joint power and subchannel allocation scheme maximizing system capacity in dense femtocell downlink systems," in *Proc. PIMRC*, pp. 1381–1385, Sept. 2009.
- [18] K. Zheng, Y. Wang, C. Lin, X. Shen, and J. Wang, "Graph-based interference coordination scheme in orthogonal frequency-division multiplexing access femtocell networks," *IET Commun.*, vol. 5, no. 17, pp. 2533–2541, Nov. 2011.
- [19] L. Li, C. Xu, and M. Tao, "Resource allocation in open access OFDMA femtocell networks," *IEEE Wireless Commun. Lett.*, vol. 1, no. 6, pp. 625–628, Dec. 2012.
- [20] C. Coletti, L. Hu, N. Huan, I. Z. Kovács, B. Vejlggaard, R. Irmer, and N. Scully, "Heterogeneous deployment to meet traffic demand in a realistic LTE urban scenario," in *Proc. VTC*, Sept. 2012.
- [21] A. Mahmud and K. A. Hamdi, "Hybrid femtocell resource allocation strategy in fractional frequency reuse," in *Proc. WCNC*, pp. 2283–2288, Apr. 2013.
- [22] G. Foschini and Z. Miljanic, "A simple distributed autonomous power control algorithm and its convergence," *IEEE Trans. Veh. Technol.*, vol. 42, no. 4, pp. 641–646, Nov. 1993.
- [23] D. Mitra, "An asynchronous distributed algorithm for power control in cellular radio systems," in *Proc. 4th WINLAB Workshop on Third Generation Wireless Info. Netw.*, 1993.
- [24] S. G. Kiani and D. Gesbert, "Maximizing the capacity of large wireless networks: Optimal and distributed solutions," in *Proc. IEEE ISIT*, pp. 2501–2505, July 2006.
- [25] S. G. Kiani, G. E. Øien, and D. Gesbert, "Maximizing multicell capacity using distributed power allocation and scheduling," in *Proc. WCNC*, pp. 1690–1694, Mar. 2007.
- [26] F. Baccelli, M. Klein, M. Lebourges, and S. Zuyev, "Stochastic geometry and architecture of communication networks," *J. Telecommun. Syst.*, vol. 7, no. 1, pp. 209–227, June 1997.
- [27] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, no. 10, pp. 1218–1230, Oct. 1997.
- [28] A. Leon-Garcia, *Probability, Statistics, and Random Processes For Electrical Engineering*, 3rd ed. Prentice Hall, 2008.
- [29] L. Narayanan, "Channel assignment and graph multicoloring," *Handbook of Wireless Networks and Mobile Computing*, pp. 71–94, 2002.
- [30] Y.-J. Chang, Z. Tao, J. Zhang, and C.-C. J. Kuo, "A graph-based approach to multi-cell OFDMA downlink resource allocation," in *Proc. Globecom*, Dec. 2008.
- [31] R. Y. Chang, Z. Tao, J. Zhang, and C.-C. J. Kuo, "A graph approach to dynamic fractional frequency reuse (FFR) in multi-cell OFDMA networks," in *Proc. ICC*, Jun. 2009.
- [32] M. Zhang, Y. Liu, M. Tao, and X. Gao, "A graph approach for coordinated channel allocation in downlink multicell OFDMA networks," in *Proc. IET ICCTA*, pp. 238–242, Oct. 2011.
- [33] B.-G. Kim, J.-A. Kwon, and J.-W. Lee, "Utility-based subchannel allocation for OFDMA femtocell networks," in *Proc. ICCCN*, Aug. 2011.
- [34] E. Pateromichelakis, M. Shariat, A. Quddus, M. Dianati, and R. Tafazolli, "Dynamic clustering framework for multi-cell scheduling in dense small cell networks," *IEEE Commun. Lett.*, vol. 17, no. 9, pp. 1802–1805, Sept. 2013.
- [35] D. Brélaz, "New methods to color the vertices of a graph," *J. Commun. of the ACM*, vol. 22, no. 4, pp. 251–256, Apr. 1979.
- [36] O. Petelin and R. Adve, "Distributed resource allocation in femtocell networks," in *Proc. CWIT*, pp. 102–107, Jun. 2013.
- [37] W. Rhee and J. M. Cioffi, "Increase in capacity of multiuser OFDM system using dynamic subchannel allocation," in *Proc. VTC*, vol. 2, pp. 1085–1089, May. 2000.
- [38] Femto Forum, "Interference management in OFDMA femtocells," *white paper*, Mar. 2010. [Online]. Available: www.smallcellforum.org
- [39] F. Hansen and F. I. Meno, "Mobile fading-Rayleigh and lognormal superimposed," *IEEE Trans. Veh. Technol.*, vol. 26, no. 4, pp. 332–335, Nov. 1977.
- [40] J. Liu, J. Wu, J. Chen, P. Wang, and J. Zhang, "Radio resource allocation in buildings with dense femtocell deployment," in *Proc. ICCCN*, Aug. 2012.



Germany. She is the recipient of the Alexander Graham Bell Canada Graduate Scholarship-Doctoral in 2010.



Raviraj S. Adve (S'88, M'97, SM'06) was born in Bombay, India. He received his B. Tech. in Electrical Engineering from IIT, Bombay, in 1990 and his Ph.D. from Syracuse University in 1996. Between 1997 and August 2000, he worked for Research Associates for Defense Conversion Inc. on contract with the Air Force Research Laboratory at Rome, NY. He joined the faculty at the University of Toronto in August 2000 where he is currently a Professor. Dr. Adve's research interests include analysis and design techniques for heterogeneous networks, energy harvesting networks and in signal processing techniques for radar and sonar systems. He received the 2009 Fred Nathanson Young Radar Engineer of the Year award.